

Durham Research Online

Deposited in DRO:

16 October 2015

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Smith, R. and Keiner, E. (2015) 'Erziehung und Wissenschaft, Erklären und Verstehen.', *Zeitschrift für Pädagogik.*, 5 . pp. 665-682.

Further information on publisher's website:

http://www.beltz.de/fachmedien/erziehungs_und_sozialwissenschaften/zeitschriften/zeitschrift_fuer_paedagogik/article/J31093cHash=916e42c5f15cfdcf7e1a32142da414a1

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Erziehung und Wissenschaft, Erklären und Verstehen¹

Richard Smith & Edwin Keiner

Klage über die Erziehungswissenschaft und erziehungswissenschaftliche Forschung ist leicht zu führen.² Sie schwebt eher in luftigen Höhen der Theorie, sie finde kaum aktuellen Anschluss an die Fortschritte anderer Wissenschaften, sie richte ihr Augenmerk nicht auf das „was wirkt“, sie lasse Theorie und Praxis unverbunden nebeneinander stehen, ihre wissenschaftlichen Standards seien ‚weich‘ etc. Die Liste der Sünden ließe sich leicht vermehren, auch wenn sie nicht alle in die gleiche Richtung weisen.

Eine Reihe von Problemen der Erziehungswissenschaft entstehen aus ihren Versuchen, medizinische Forschungskulturen zu imitieren; dies gilt insbesondere für randomisierte kontrollierte Studien (randomised controlled trials = RCTs), die als ‚Goldstandard‘ in der medizinischen Forschung betrachtet werden. Studien dieser Art werden dafür genutzt zu zeigen „was wirkt“, und ihre Relevanz für pädagogische Kontexte wird zunehmend als selbstverständlich betrachtet (Smeyers 2006).³ Als ebenso selbstverständlich gilt, dass ‚wissenschaftliche‘ Forschung spezifischen Kriterien von Strenge und Robustheit genügen solle, ohne dass dies selbst noch einmal kritisch bestimmt werden würde. Darüber hinaus wird davon ausgegangen, als ‚Forschung‘ im eigentlichen Sinne gelte stets empirische Forschung, die auf Gewissheit (Keiner 2006) nach Art der Naturwissenschaften ziele. (Smeyers 2007).⁴ Nach diesem Konzept lernt der gute Forscher zunächst wie man ein Forschungsprojekt plant und durchführt (Hodgson & Standish 2007) und geht dann zügig ins ‚reale‘ Feld, um Daten zu sammeln.

Erziehungswissenschaftliche, Empiri(sti)sche, ‚evidence based‘ Forschung

Die Leuener Forschungsgruppe hat diskutiert, ob und inwiefern eine gewisse Rhetorik, Metaphorik und Sprachspiele unvermeidlich zur Forschung gehören, und ob man dies nicht eher als Spezifikum denn als Schwäche betrachten müsse (Smeyers & Depaepe 2010). Manche der in der Wissenschaftlergemeinschaft besonders geschätzten Forschungsmethoden und -techniken, so wurde argumentiert, müssten ergänzt werden um Methoden etwa der Geschichtswissenschaft oder der Philosophie (Smeyers & Depaepe 2010). Inwiefern, so wurde diskutiert, werde Erziehungswissenschaft gegenwärtig im Interesse an ‚Reformen‘ diszipliniert und kontrolliert (Smeyers & Depaepe 2008). Und sie stellte fest, dass das, was auf der Grundlage eines quasi-

¹ Der Text wurde im Rahmen der Leuener Forschungsgruppe diskutiert und aus dem Englischen ins Deutsche übersetzt. Wir beziehen uns in Teilen auf Smeyers & Smith: *Understanding Education and Educational Research*, das demnächst bei Cambridge University Press erscheinen wird.

² Als zwei bekannte kritische Positionen aus dem angloamerikanischen Bereich s. Hargreaves 1996; Tooley 1998; für den deutschsprachigen Bereich s. etwa Tenorth 1983, Hoffmann & Neumann 1998; zur aktuellen Reflexion Fatke & Oelkers 2014.

Im Folgenden übersetzen wir das englische „educational research“ mit Erziehungswissenschaft. Wir wissen, dass wir damit unterschiedliche Bedeutungen von Begriffen zwischen, aber auch innerhalb unterschiedlicher Wissenschaftskulturen ignorieren – man denke nur an die Unterschied zwischen education research, educational research und educational studies oder von Pädagogik, Erziehungswissenschaft, Bildungswissenschaft, Bildungsforschung (vgl. Keiner 2015). Eine solche Analyse muss als Desideratum markiert werden; sie ist hier nicht zu leisten.

Schließlich bemerken wir, dass wir – unabhängig von dem gewählten Genus – stets das männliche und weibliche Geschlecht gleichermaßen meinen.

³ Solche Annahmen und Voraussetzungen sind ausführlich von der Leuener Forschungsgruppe diskutiert und in ihren Ergebnissen 2006 in einem Buch mit dem Untertitel „Why ‚What Works‘ Doesn’t Work“ veröffentlicht worden (Smeyers & Depaepe 2006).

⁴ Mit diesem Thema hat die Leuener Forschungsgruppe ihre Arbeit begonnen; ihr erstes Buch trägt den Titel ‚Beyond Empiricism‘ (Smeyers & Depaepe 2003).

naturwissenschaftlichen Verständnisses erforscht werden kann, nicht notwendig mit dem übereinstimmt, was als pädagogisch sinnvoll und wertvoll erachtet wird – auch dass z.B. Lernen und Erziehung nicht dasselbe sind (Smeyers & Depaepe 2013, Burbules 2013).

Nehmen wir ein Beispiel empirischer Forschung: eine randomisierte kontrollierte Studie, die im Vergleich die Wirksamkeit zweier Ansätze synthetischer Phonetik untersucht. Es werden zwei Klassen der gleichen Schule ausgewählt. Da die Kinder die gleiche soziale Zusammensetzung und den gleichen Bildungslevel aufweisen, wird angenommen dass diese beiden Klassen auch in weiteren Hinsichten identisch seien – eine Annahme, die meist zu schnell getroffen wird (Cartwright & Hardie 2012). Die beiden Lehrer wurden ausgebildet, die Kinder in gleicher Weise zu unterrichten – eine weitere fragwürdige Annahme. Was werden die Tests am Ende ergeben? Sie können nur zeigen, dass im Test die Kinder in Klasse A besser abgeschnitten haben als diejenigen in Klasse B. Ob der eine Ansatz in pädagogischer Hinsicht sinnvoller ist als der andere, würde aber bedeuten, dass man z. B. zeigen könnte, dass die Kinder zu begeisterten und anspruchsvollen Lesern geworden seien: Sie fangen an, sich Bücher als Geburtstagsgeschenke zu wünschen, sie besuchen die örtliche Bücherei, sie interessieren sich für Buchbesprechungen interessieren und kaufen und lesen Romane, Biografien, Gedichte etc.. Dies zu erforschen würde jedoch einen erheblich längeren Forschungszeitraum erfordern als die Geldgeber akzeptieren würden. Darüber hinaus würden im Zeitraum zwischen dem Einsatz der neuen Lehrtechnik und dem Ergebnis, ein anspruchsvoller Leser geworden zu sein, viele intervenierende Variablen auftreten, die es schwierig, wenn nicht gar unmöglich machen, eine eindeutige kausale Schlussfolgerung zu ziehen.

Die Annahme, man könne die Welt in direktem, empiristischem Zugang erschließen, zeigt sich auch im akademischen Alltagsleben, zumindest in Großbritannien. Im Folgenden präsentieren wir beispielhaft einige wahre (anonymisierte) Begebenheiten.

Eine Bewerberin für eine Professorenstelle im Bereich der Sozialwissenschaften (nicht Erziehungswissenschaft) an einer berühmten britischen Universität wurde von der Vorsitzenden der Berufungskommission gefragt: „Jetzt haben sie uns alles über ihre Ideen und Theorien berichtet; wie sieht es mit ihrer gegenwärtigen Forschung aus?“ Scheinbar ist ‚echte‘ Forschung etwas anderes als denken und schreiben; es ist offensichtlich etwas, wozu man hinausgehen und etwas finden muss. Forschung hat empirisch zu sein oder sie ist nicht im eigentlichen Sinne Forschung.

Eine andere Berufungskommission: die Bewerberin, die u.a. mehrere Analysen von Bildungspolitik veröffentlicht hat, wurde gefragt, in welcher Weise diese Arbeiten als ‚Forschung‘ gezählt werden könnten. Ihre Arbeiten stützten sich nicht auf harte Daten und könnten deshalb nur als Meinungsäußerung betrachtet werden.

Eine Studentin im letzten Jahr vor Abschluss ihres Studiums wollte sich in ihrer Abschlussarbeit mit dem Thema ‚Selbstwertgefühl und internationale Pfadfinderinnenbewegung‘ beschäftigen. Sie selbst war eine begeisterte Pfadfinderin und Gruppenleiterin und hoffte herauszufinden, dass Mitarbeit bei Pfadfinderinnen das Selbstwertgefühl steigere. Sie beabsichtigte eine repräsentative Anzahl von Pfadfinderinnen zu befragen, ob sie der Auffassung seien, dass die Mitgliedschaft bei den Pfadfinderinnen ihr Selbstwertgefühl gesteigert hätte. Auf einer Skala von 1-10 sollten sie angeben, welchen Level von Selbstwertgefühls sie vor Eintritt in die Organisation gehabt und welches sie jetzt hätten, und welche Aktivitäten ihr Selbstwertgefühl verbessert hätten, usw. Es geht hier nicht darum, diese Studentin zu kritisieren. Es war jedoch nicht einfach, sie zu überzeugen, dass die Dinge nicht ganz so einfach lägen. Das gilt zunächst für die Feststellung von Ursachen; es ist schwierig zu zeigen, dass das, von dem jemand meint, es habe sein Selbstwertgefühl verbessert, auch in der Tat diese Verbesserung verursacht hat. Weit schwieriger und komplexer ist freilich das Konzept Selbstwertgefühl selbst. Die Studentin hatte ein Pflichtmodul über Forschungsmethoden besucht, in dem sie gelernt hatte, dass gute Forschung mit begrifflichen Problemen anfängt, die man durch klare Definitionen, Handbücher und Lexika löst und dass dann ‚echte‘ Forschung in der Durchführung von

Interviews und Fragebögen, im Codieren und Auswerten und im Schlussfolgern valider Aussagen nach folgendem Muster bestünde: „67% meiner Befragten stimmten zu, dass die Mitarbeit bei Pfadfinderinnen ihr Selbstwertgefühl verbessert hat. Vollständig ausgefüllte Fragebögen wurden jedoch nur von neun Personen zurückgegeben, das sind weniger als 20% der versendeten Fragebögen. Vor diesem Hintergrund sei es nötig, weitere Forschung auf diese wichtige Frage zu richten. (Die Sätze sind mehreren unterschiedlichen, aber ähnlichen Abschlussarbeiten entnommen).

Eine frisch berufene Professorin wurde eingeladen, an einem Gemeinschaftsprojekt zwischen ihrer neuen Universität und einer Universität in den USA teilzunehmen – eine Ehre, so wurde ihr gesagt –, um „ein umfangreiches Evaluationsinstrumentarium zu entwickeln, das in der Lage sei, hochbegabte Kinder zu identifizieren und zu fördern“. Ihre Aufgabe bestand darin, Eltern zu interviewen, die ihre Kinder in ein Programm zu Diagnose und Entwicklung von Hochbegabung gegeben hatten, um herauszufinden, auf welcher Grundlage diese Eltern ihre Kinder als hochbegabt einschätzen. Letztendlich zog sich die Wissenschaftlerin aus dem Projekt zurück, hauptsächlich deswegen, weil ihrer Auffassung nach der Sachverhalt von Hochbegabung unbefragt vorausgesetzt wurde und es nur darum ging, wie Hochbegabung diagnostiziert werden könne. Kulturelle, soziologische oder philosophische Aspekte blieben unberücksichtigt. Aus diesen Gründen war ihr auch die technologische Vorstellung eines ‚Evaluationsinstrumentariums‘ zuwider. Ihr Rückzug wurde mit Bestürzung registriert. Gab es nicht eine erhebliche Menge an Drittmitteln für dieses Projekt, das sich doch gut in ihrer Vita und Publikationsliste auswirken könnte? Ihr Anliegen einer Berücksichtigung auch kultureller und philosophischer Fragen führte schließlich zu der sarkastischen Zuschreibung, dass sie offensichtlich zu einer Sorte von Menschen gehöre, die sich lieber in einer Bibliothek aufhalte als sich ‚ihre Hände schmutzig zu machen‘.

In diesem Beispiel, und viele weitere könnte man hinzufügen, geht es recht direkt um die Menge an eingeworbenen Drittmitteln und die Vermittlung des Eindrucks transatlantischer Geschäftigkeit, verbunden mit wenig Reflexivität und Zweifel an einem allgemeinen Selbstverständnis einer Wissenschaftlichkeit, die in einer spezifischen Art einer Rhetorik der Bedeutsamkeit verankert ist. Nüchtern betrachtet erfolgt erziehungswissenschaftliche Forschung – ebenso wie viele andere Formen wissenschaftlicher Forschung auch – in spezifischen Materialien und realen Kontexten, die zumindest teilweise die Themen, den Stil unseres Schreibens, den disziplinären Hintergrund, von dem aus wir denken etc., bestimmen.

Gegenwärtig erfährt eine spezifische Rhetorik der Bedeutsamkeit besondere Aufmerksamkeit: Zahlen und Schaubilder, Tabellen und statistische Maßzahlen. Die Symbolsprache statistischer Konzepte ist gegründet in Formaldefinitionen jenseits der gesprochenen Sprache. Da diese statistische Sprache innerhalb ihres eigenen symbolischen Formats schwierig zu lehren und zu verstehen ist, wird die gesprochene Sprache genutzt, um auf nicht-formale Weise statistische Konzepte zu definieren. Innerhalb dieser nicht-formalen Definitionen wird die statistische Symbolsprache ersetzt durch geeignete Wörter der gesprochenen Sprache. Dies bedeutet, dass Analogien und Metaphern, die viele unterschiedliche Bedeutungen in der Alltagssprache annehmen können, die Konstruktion von statistischer Bedeutung in hohem Maße bestimmen. Dies bedeutet auch, dass wir, wenn wir methodologische Vorannahmen solcher quantitativen Forschung verstehen wollen, uns zuweilen in die Rolle von Literaturkritikern begeben sollten, um Sinn und Bedeutung hinter den Wörtern und Metaphern zu finden (Stadler Altmann & Keiner 2010).

In Großbritannien dominiert eine formale, regelmäßig wiederkehrende Evaluation wissenschaftlicher Forschung, die seit über 20 Jahren unter der Federführung unterschiedlicher nationaler ‚Research Assessment Exercises‘ (RAEs) durchgeführt wurde und jetzt unter dem Titel ‚Research Excellence Framework‘ (REF) agiert. Es gibt zahlreiche Aspekte des REF, die die Tendenz zu einem unreflektierten Empirismus und Scientismus zeigen können. In den Evaluationsprogrammen wird der

Gegenstandsbereich von Erziehung behandelt als wäre er ein Teil der Psychologie oder der Medizin. Die Tatsache, dass sozialwissenschaftliche Fachzeitschriften es mittlerweile als selbstverständlich erachten, dass eingereichte Beiträge den Zitationsregeln und Konventionen der American Psychological Association (APA) folgen (Burbules 2013), ist sowohl Ursache wie Folge in einem iterativen Prozess. Zeitschriftenartikel wiederum werden als qualitativ hochwertiger angesehen als Buchbeiträge oder sogar ganze Bücher, weil in der Psychologie im allgemeinen davon ausgegangen wird, dass Spitzenforschung in der neuesten Ausgabe einer prominenten Fachzeitschrift erscheinen wird, wohingegen ein Buch im Wesentlichen als ein Hand- oder Lehrbuch gilt, das eher den gegenwärtigen Stand der Forschung zusammenzufasst als die neuesten Forschungsergebnisse verbreitet. Eine neue Entdeckung kann nicht so lange warten, so die Annahme, bis der aufwändige Prozess der Publikation eines Buches abgeschlossen ist. (In der Erziehungswissenschaft wird man freilich auch die Frage stellen müssen, ob es solche vergleichbar ‚dramatischen‘ Forschungsergebnisse überhaupt gibt.) Ein weitere Folge ist, dass Forscher ermutigt werden, sich bereits etablierten Forschungsteams anzuschließen, um damit, wie in der Psychologie, zu einem von vielen Autoren eines Zeitschriftenartikels zu werden – als ob erziehungswissenschaftliche Forschung kumulativ und nach Art eines Puzzles funktioniere. Man wird auch berücksichtigen müssen, dass die Psychologie als eine wissenschaftliche Disziplin darauf zielt, das ‚Subjekt‘ zugunsten einer naturwissenschaftlichen Betrachtungsweise weitgehend zu eliminieren – was sich manchmal in erregten, wenngleich nicht notwendigerweise realistischen Erwartungen äußert, dass ein tieferes und komplexeres Verstehen von Menschen über die Analyse der DNA, durch die Hirnforschung oder die Neurobiologie zu erreichen sei. Spezialisten dieser Bereiche findet man nun mehr und mehr in erziehungswissenschaftlichen Departments an Universitäten. Dies gibt uns den Anlass zu bemerken, dass Erziehungswissenschaft, will sie nicht ‚fremd‘ disziplinär vereinnahmt werden, gut daran täte, eine eigene Stimme und Perspektive zu entwickeln.

Die Annäherung der Erziehungswissenschaft an die Medizin wird deutlich in der sich erstaunlich schnell verbreitenden Erwartung, dass randomisierte kontrollierte Studien (RCTs) in der Erziehungswissenschaft die gleiche Rolle spielen könnten wie in der Medizin. Das britische Erziehungsministerium veröffentlichte am 3. Mai 2013 eine Presseerklärung mit dem rhetorisch suggestiven Titel „New randomised control trials will drive forward evidence based research“. Diese Initiative wurde nachdrücklich unterstützt, vielleicht sogar weiter vorangetrieben, von dem Wissenschaftsjournalisten Ben Goldacre. In seiner Serie ‚Schlechte Wissenschaft‘ in The Guardian, 14. Mai 2011, schreibt er:

Randomised trials are our best way to find out if something works: by randomly assigning participants to one intervention or another, and measuring the outcome we're interested in, we exclude all alternative explanations for any difference between the two groups. If you don't know which of two reasonable interventions is best, and you want to find out, a trial will tell you.

Und Goldacre zweifelt nicht daran, dass solche randomisierten kontrollierten Studien ausgeweitet und übertragen werden können auf Sozialpolitik, insbesondere auf den Bereich der Bildung und Erziehung.

Do free uniforms improve school attendance, especially in pupils who don't own one at all? ... Do long prison sentences work? At the moment, sentences are hugely variable anyway: randomise properly and run a trial. Different teaching approaches? Run a trial. Harder exams? Run a trial. Job-seeking support? Run a trial.

Trotz der suggestiven Rhetorik gibt es jedoch zumindest zwei Gründe zur Skepsis gegenüber solchen und ähnlichen Erwartungen an RCTs. Der erste besteht in dem Problem der voneinander abhängigen, konfundierten Variablen. Man wird davon ausgehen können, dass die Einführung kostenloser Schuluniformen andere Effekte haben wird, wenn Eltern dies begeistert unterstützen, als wenn dies

nicht erfolgt. Eine Politik, die den Außenseiterstatus einiger Kinder zu verhindern sucht, wird andere Effekte haben als eine, die auf eine große Minderheit von Schülerinnen und Schülern gerichtet ist, die ihre Identität durch Widerstand gegen Schulnormen und -regeln gewinnen und definieren. Eine Schuluniform-Politik, die in Kenia funktioniert (wohin ein [Hyperlink](#) in der online-Version des Artikels von Goldacre verweist), dürfte wohl kaum in Manchester funktionieren, weil die kulturellen Kontexte so unterschiedlich sind. Ähnliches gilt für unterschiedliche Lehrmethoden; in der online Version des Artikels verweist ein [Hyperlink](#) auf die Debatte ob es sinnvoll sei, Kindern das Lesen mithilfe eines phonetischen Ansatzes beizubringen. Ein Lehrer, der sorgfältig, ja ängstlich den Vorschriften und Regeln eines solchen Ansatzes folgt, unterrichtet anders als ein Lehrer, der sich entspannt und flexibel verhält, selbst wenn es schwierig sein sollte signifikante Unterschiede ihrer Unterrichtsmethode zu finden.

Der zweite Grund zur Skepsis ist die Tatsache, dass es bei einem medizinischen RCT nachvollziehbar und erklärbar ist, ob eine bestimmte Intervention funktioniert: der Blutdruck sinkt bei der einen Gruppe von Patienten; weniger Patienten klagen über Migräne. Dies gilt jedoch nicht im Bereich der Sozial- und Bildungspolitik. Ob die neue Einführung von Schuluniformen, als Erfolg bewertet werden kann hängt davon ab, welche Kriterien man für Erfolg wählt. Als Kriterien könnten gelten: Schulbesuchsquote und Bildungsbeteiligung, Prüfungsergebnisse, Elternwahlverhalten und Anmeldungen, Verbleibquoten jenseits der Pflichtschulzeit, Verringerung von Mobbing, Anzahl von Konflikten mit Schülern aus Nachbarschulen etc.. Wenn Jugendliche zu längeren Haftstrafen im Gefängnis verurteilt werden, mag der Erfolg darin bestehen, dass sie sich nicht mehr ‚auf der Straße herumtreiben‘. Folgeeffekte dürften aber kaum als Erfolge zählen, wenn sie nach längerer Haftstrafe wieder in die Gesellschaft entlassen werden – brutalisiert und entfremdet sowie fähig und bereit zu weiteren Straftaten, über die sie sich während der langen Haftzeit ‚weiterbilden‘ konnten. Zu prüfen wäre auch die Frage, inwiefern RCTs selbst zu Umfang und Höhe der Urteilsprüche beitragen.

Vor einigen Jahren haben Cartwright & Hardie (2012) in einem Buch über Evidence-based Policy festgestellt, dass RCTs uns in der Tat nicht sagen, ‚was wirkt‘, sondern nur sagen können, dass eine bestimmte Maßnahme in einem bestimmten Zeitraum in einer bestimmten Schule, Krankenhaus oder einem anderen Setting gewirkt hat. Um solche Ergebnisse generalisieren zu können müssen wir wissen, was Schulen oder Krankenhäuser im Wesentlichen ausmacht, wie wir sie vergleichen können usw., und dies erfordert ein Urteil oder, wie Cartwright & Hardie es ausdrücken: Beratung und Erwägung (deliberation). Sie geben ein gutes Beispiel darüber, wie ein Ernährungsprojekt in Tamil Nadu, Indien, das besonders darauf ausgerichtet war, schwangeren Müttern zu zeigen, wie sie ihre Kinder ernähren sollen, in Bangladesch nicht funktionierte – trotz scheinbar ähnlicher Kontexte. Es stellte sich heraus, dass in ländlichen Bereichen von Bangladesch die Mütter nicht immer den Einkauf der Nahrung kontrollieren konnten, weil dort die Männer einkaufen gehen. Und sie hatten auch keine Kontrolle über das, was ‚auf den Teller‘ kam, weil eine signifikante Minderheit von Frauen mit ihren Schwiegermüttern zusammen lebte, denen gegenüber sie sozial untergeordnet waren. Beratung und Erwägung ist also notwendig, um beantworten zu können, ob das, was als ‚wirksam‘ identifiziert wird, als Kausalverhältnis zu verstehen sei. Cartwright & Hardie (2012) schreiben unter Bezug auf Aristoteles, dass die herrschende Lehre ein System von Regeln sei, d.h., evidence based policy wird im Wesentlichen wahrgenommen als eine Angelegenheit, bei der Regeln und Prozeduren, die in der Schule oder dem Krankenhaus X funktioniert haben, auf die Schule oder das Krankenhaus Y übertragen werden. Dieses, so bemerken sie

discourages decision makers from thinking about their problems, because the aim of rules is to reduce or eliminate the use of discretion and judgement, and deliberation requires discretion and judgement. The aim of reducing discretion comes from a lack of trust in the ability of operatives to exercise discretion well. Whether it is possible to reduce discretion depends on whether the process of deciding what will be effective...can be reduced to the operation of rules. We say that it often, or typically, cannot. And that if it cannot, the attempt to replace discretion with rules, such as ‘Do it if, or maybe

only if, it has worked there', is very damaging. Deliberation is not second best, it is what you have to do, and it is not *faute de mieux* because there is no *mieux*. (p. 158)

An anderer Stelle notieren Sie, dass "mandating RCT-based policies selects in favour of operatives who are good at conforming with rules and against those who are good at thinking" (p. 11). Dies bedeutet auch, dass diese hegemoniale Form erziehungswissenschaftlicher Forschung uns nicht nur nicht sagt, 'was wirkt'; sie leiht darüber hinaus ihr 'wissenschaftliches' Gewicht denjenigen Kräften, die dazu beitragen, den interpretierenden, abwägenden und urteilenden Blick zu verengen.

Der zunehmende Einfluss von RCTs ist nur ein Element unter einer Vielzahl anderer Faktoren, die zu etwas geführt haben, was wir die Autonomie von quasi-wissenschaftlichen Techniken in der Forschung nennen könnten. Studierende der Physik, Biologie oder Chemie müssen Labortechniken lernen – einen Oszillator oder ein Mikroskop zu bedienen, Kulturen in Petrischalen anzulegen oder Kristallwachstum und Schmelzpunkte zu bestimmen. Es ist insofern nur ein kleiner Schritt zu der Vorstellung, dass jedes Wissen und jedes Verständnis den gleichen Zugang benötige. Dies hängt eng zusammen mit der uralten Annahme, es gäbe so etwas wie 'Forschungsmethoden', die man sich brav aneignen könne. Diese wurde, gerade auch für die Erziehungswissenschaft, weiter genährt durch mobile, ausländische Studierende, deren sprachliche Kompetenzen beeindruckend gut sind, die sie aber als Nicht-Muttersprachler kaum dazu zu befähigen, sich mit begrifflichen oder philosophischen Dimensionen von Erziehung und Erziehungswissenschaft angemessen zu beschäftigen.

Unterschiedliche Techniken des Datensammelns für empirische Forschungsprojekte sind demgegenüber relativ einfach zu verstehen, und sie bieten Entlastung: je schneller man zu Zahlen und Quantitäten kommt, desto schneller kann man die verzwickten Probleme von Begriffen und philosophischen Vorstellungen hinter sich lassen. T-test, ANOVA, ANCOVA, Chi-Quadrat, Lineare Regression, Faktorenanalyse und eine Vielzahl weiterer statistischer Techniken haben darüber hinaus den Vorteil, Erziehungswissenschaft als 'wissenschaftlich' aussehen zu lassen. Diese Techniken sind insofern 'autonom', als sie unterstellen, man könne sie unabhängig von einem spezifischen Gegenstands- oder Untersuchungsbereich lernen und dann auf eine große Bandbreite sozialwissenschaftlicher Projekte übertragen, gleich ob sie sich dem Namen nach mit Erziehung beschäftigen oder mit neuen Technologien oder dem Drogenmissbrauch.

Dies Phänomen findet sich nicht nur in Großbritannien oder Deutschland. Wir beobachten einen europäischen, ja sogar weltweiten Prozess der Standardisierung dessen, was als gute, 'wissenschaftliche' Forschung zählt, und eine Instrumentalisierung von Forschungswissen zu einem politisch und ökonomisch nützlichen Wissen, das zur Entwicklung der sogenannten 'Wissensgesellschaft' beitragen soll.

Obwohl diese Trends der Erziehungswissenschaft gegenwärtig sehr dominant und mit externen Erwartungen verbunden sind, lassen sich Kritik und Skepsis auch an die internen disziplinären Muster der Problembearbeitung adressieren, auch an diejenigen Fraktionen der Erziehungswissenschaft, die diese Trends aus einer kritischen Perspektive untersuchen, analysieren und reflektieren. Kritik und Skepsis lassen sich auf zwei Aspekte ausrichten: a) auf den Beitrag, den erziehungswissenschaftliche Forschung selbst zu solchen Trends und zu ihren kulturellen Diversifizierungen leistet, und b) auf die Bedeutung von empirischer, historischer und vergleichender Analyse und Reflexion solcher Entwicklungen, von Forschung über erziehungswissenschaftliche Forschung, kurz: von Wissenschaftsforschung.

Zu a) gegenwärtig haben Rhetoriken Konjunktur, die sich auf zukunftsbezogene Erwartungen an lebenslanges Lernen, 'competition and cooperation', selbst-konstruierte Konstitutionen des Subjekts etc. beziehen. Die 'Wissensgesellschaft' ist eine solche moderne Ideologie, eine universale Erwartung die sich an alle Individuen richtet und sie auffordert zu lernen, zusammenzuarbeiten, in Wettbewerb

zu treten und schließlich sich verantwortlich zu fühlen für die Konstruktion ihres eigenen Lebenslaufs, obwohl sie nicht über die Markierungen und Kriterien eines erfolgreichen Lebens verfügen. Erziehung- und Bildungsinstitutionen fungieren als Lebenslaufregimes, die ihr Klientel mit solchen Markierungen überziehen und dabei – unterstützt durch eine groß angelegte administrative und Forschungsmaschinerie – Massen standardisierter Daten sammeln, um den Erfolg der Institution und ihrer Klienten nachzuweisen. Diese Maschinerie ist darauf angelegt, über Kriterien zu verhandeln und zu beschließen, Kategorien zu definieren, Daten zu sammeln, auszuwerten und grafisch darzustellen – ein gigantisches Unternehmen, das sowohl der Aufklärung ihre Krone auf- und dem Streben nach Glückseligkeit ein Ende setzt. Der Satz, ‚gute Entscheidungen sind informierte Entscheidungen‘ fungiert als eines der großen Mantras moderner Rationalität, und von Sozial- und Erziehungswissenschaften wird selbstverständlich erwartet, dass sie einen signifikanten Beitrag dazu leistet. Ökonomische, soziale und pädagogische Fortschrittserwartungen geben die normativen Bezugspunkte ab, die den Bedarf an standardisierter, ‚evidence-based‘ und vergleichbarer Forschung vorantreiben.

In gewisser Weise trägt die Erziehungswissenschaft selbst zu solchen Dynamiken bei, zum einen durch ihre Bereitschaft, relativ schnell auf externe Themen und Erwartungen zu reagieren, ohne sie zu brechen und nach einer systematischen, disziplinären Logik umzudefinieren (Bourdieu 1998, p. 19). Ob solche Themen im Modus erziehungswissenschaftlicher Forschung oder im Modus reformorientierter Reflexion aufgenommen werden, sind nur zwei Seiten der gleichen Medaille. Zum anderen inkorporiert die Erziehungswissenschaft eine breite Palette normativer Orientierungen, ablesbar etwa an Diskussionen um Ziele und Mittel, Reformen, Aufgaben, Herausforderungen oder an den endlosen Debatten über das Theorie-Praxis-Verhältnis. Sie tendiert zu einer Pädagogisierung sozialer Probleme (Smeyers & Depaepe 2008), selbst im Modus erziehungswissenschaftlicher Forschung (Keiner 2006) und ist zuweilen sogar dazu verführbar, sich in normativer Selbstüberschätzung an politischer Macht beteiligen zu wollen.

Solche disziplinären Muster der Kommunikation variieren freilich entsprechend unterschiedlicher Forschungskulturen. Sie ergeben sich aus unterschiedlichen Forschungsorientierungen, Methodenpräferenzen, theoretischen Schlüsselkonzepten, Bezügen zu Nachbardisziplinen, Konzeptualisierungen des Theorie-Praxis-Verhältnisses, Formen des transnationalen Wissensaustauschs etc., aber auch aus unterschiedlichen Organisationsformen und infrastrukturellen Bedingungen (Keiner & Schriewer 2000).

Zu b) Die Bezüge und Kombinationen der unterschiedlichen disziplinären Muster erziehungswissenschaftlicher Kommunikation in Bildungsphilosophie und Bildungsgeschichte – inklusive ihrer unterschiedlichen methodologischen Bezugshorizonte – waren seit Anbeginn im Fokus der Leuener Forschungsgruppe. Mit einem solchen Blick gewinnen Fragen nach dem intellektuellen Esprit und interdisziplinärem Denken, sowie zur Selbstreflexion und Selbststeuerung wissenschaftlicher Gemeinschaften an Bedeutung. Novoa & Yariv-Mashal (2003) unterscheiden kritisch – an Beispiel der Vergleichenden Erziehungswissenschaft – zwischen Analysen mit dem Interesse an gesellschaftlicher Steuerung und Analysen mit dem Interesse an intellektuellen Bildungsreisen. Diese Unterscheidung kann leicht auf empirische Forschung (etwa zwischen evidence-based und methodisch spielerischen Ansätzen) oder historische Forschung (etwa zwischen traditionalisierenden und an Irritationen interessierten Ansätzen) in der Erziehungswissenschaft übertragen werden. Diese Unterscheidung enthält jedoch pejorative Konnotationen: Erstere zeuge von schlechtem Geschmack und sei mit politischen ‚Schmutz‘ kontaminiert, die letztere sei demgegenüber eine unschätzbare, hochgebildete intellektuelle Tätigkeit, die denjenigen Habitus kennzeichne, den man von einem richtigen Forscher eigentlich erwarten müsse. Verstünde man aber das ‚Interesse an gesellschaftlicher Steuerung‘ als Interesse an wissenschaftlicher Selbststeuerung, dann würden die pejorativen Konnotationen ihr Gewicht verlieren. Alle Diskussionen über Forschungsqualität beziehen sich ja auch auf Selbststeuerung. Insofern sind Geschichte, Philosophie, empirische

Forschung und Vergleich nicht alleine als intellektuelle Bildungsreisen, sondern auch als Medien der Selbstreflexion und -steuerung zu betrachten. Sie dienen auch als Rahmen, Form, Landkarte, die uns orientieren und nach denen wir uns explizit entscheiden etwas zu untersuchen und zu ‚reisen‘ – bevorzugt in intellektuell reiche und herausfordernde Regionen.

Ein erstes Beispiel für eine mögliche disziplinäre Selbstaufklärung ergibt sich aus einer kleinen empirischen und vergleichenden Analyse von Zeitschriften der historischen Bildungsforschung (Keiner 2013). Historische Bildungsforschung scheint gegenwärtig Themen der Zeitgeschichte besonders zu präferieren. Die Artikel in den neueren Ausgaben dieser Zeitschriften beziehen sich – anders als ältere – weitgehend auf Geschichte des 20. und sogar 21. Jahrhunderts. Diese Wendung zu einer solchen Art Präsentismus könnte auch einen disziplinären Wandel von einer disziplinär eher geschichtswissenschaftlich bestimmten zu einer eher erziehungswissenschaftlich bestimmten historischen Bildungsforschung indizieren. Was bedeutet ein solcher Befund für das Fach? Gibt es Notwendigkeiten der Selbststeuerung? In welche Richtung?

Ein zweites Beispiel bezieht sich auf unterschiedliche disziplinäre Forschungskulturen, -stile und Denkformen und deren Folgen. Meusburger (2009) und Ambrose (2006) unterscheiden ‚fractured-porous disciplines‘ und ‚unified-insular disciplines‘. Wir nutzten diese Unterscheidung für eine vergleichende Analyse erziehungswissenschaftlicher Fachzeitschriften und Fachzeitschriften der pädagogischen Psychologie (Knaupp et al. 2014; Keiner & Schaufler 2014). Erziehungswissenschaft, so lässt sich zeigen, weist Charakteristika der ‚fractured-porous disciplines‘ auf: interner Dissens über Theorien und Methoden, schwache disziplinäre Abgrenzung, negative Importbilanz von Wissen, aber hohe Kreativität und innovatives Potenzial. Pädagogische Psychologie lässt sich demgegenüber den ‚unified-insular disciplines‘ zurechnen: hoher interner Konsens über Theorien, Methoden, Forschungsstandards und Evaluationskriterien, klare und strikte disziplinäre Abgrenzung, wenig interdisziplinärer Austausch, ‚normale Wissenschaft‘ und wenig Kreativität. Was bedeuten eine solche Unterscheidung und ein solcher Befund? Gibt es Notwendigkeiten der Selbststeuerung? In welche Richtung?

Engagement und Distanzierung, Erklären und Verstehen

Zweifellos hat der Zauber der ‚Wissenschaft‘ eine Menge Antworten auf all diese vielen Fragen. Ein ‚wissenschaftliches Zeitalter‘ ist ohne Frage offensichtlich besser als ein unwissenschaftliches. Oancea & Pring (2008, p. 25) können sehr schön zeigen, dass Politiker beim verzweifelten Versuch, Gewissheit und sicheren Grund für ihre politische Linie zu finden, wissenschaftliche Hinweise und Indizien (evidence) mit einem Beweis (proof) verwechseln. Als Beweis wird etwas gesehen, das unbedingte, zwingende Überzeugungskraft aufweist, das den Politiker seiner Verantwortung enthebt, wenn etwas schief geht. Es hat darüber hinaus den Vorteil, den Politiker davon zu entlasten, sein eigenes Urteilsvermögen gebrauchen zu müssen. Urteile, so Oancea & Pring, bereiten gerade solche Personen Sorgen, die zur Annahme neigen, dass es nur wenig Platz zwischen einem wissenschaftlichen Beweis auf der einen und bloßer Subjektivität auf der anderen Seite gibt.

Gerade dieser ‚schmalen Spalt‘ aber ist reich und komplex. Dies lässt sich durch einen kurzen Blick auf die klassische und wichtige Unterscheidung zwischen Erklären und Verstehen (Wright 1971) zeigen – das eine näher an den Naturwissenschaften, das andere weniger. Ein einfaches Beispiel für das, was unter Erklärung gemeint ist (vgl. Midgley 1996): man entdeckt eine kleine Wasserpfütze im Wohnzimmer und sucht herauszufinden woher sie kommt: es könnte an einem Leck in der Rohrleitung unter dem Dielenboden der oberen Etage liegen, Wasser könnte aus einem Abflussrohr heraus laufen, jemand könnte zu sorglos seine Topfpflanze gegossen haben oder die Katze hat ihr Wasser nicht halten können. Wenn die Wasserpfütze auf eine diese Ursachen zurückgeführt werden

kann, wird man die anderen ausschließen können. Die Untersuchung folgt dem Muster einer Erklärung, indem man Licht auf den Sachverhalt wirft und ihn in wörtlichem Sinne ‚aufklärt‘. Eine empirische Untersuchung ist in diesem Fall ein richtiger und angemessener Ansatz, wenn man ausreichend Licht hat, um zu sehen, was geschehen ist.

Im Vergleich dazu nehmen wir an, eine Kollegin zeige bei Sitzungen nur einen frostigen Gesichtsausdruck, ihre Studenten klagen darüber, sie sei langweilig, mehrmals findet man sie an ihrem Schreibtisch mit dem Kopf in ihren Händen. Mit gutem Grund könnte man darüber nachdenken ob sie an Depressionen leidet, ob sie ihre berufliche Motivation verloren hat oder ob sie Symptome von Überarbeitung zeigt. Es könnte freilich auch an einer Kombination aller dieser und weiterer Faktoren liegen. Es gibt keine klare Unterscheidung zwischen Depression und fehlender Motivation. Es werden keine investigativen Techniken zur Anwendung gebracht, von denen man erwarten könnte, dass man an irgendeinem Punkt eine ‚richtige Antwort‘ finden würde. Wir bewegen uns im Bereich des Verstehens. Man benutzt sein normales menschliches Verstehen, um zu versuchen, dem Zustand und Verhalten der Kollegin Bedeutung und Sinn zu geben. Dafür braucht man ein gewisses Verständnis darüber, was es bedeutet, sich depressiv zu fühlen. Insofern ist es hilfreich, dass die meisten Menschen im Verlaufe ihres Lebens schon einmal unglücklich gewesen sind, wenn nicht gar ‚klinisch depressiv‘; und zugleich muss man die Vorzustellung vermeiden, das Leben der Kollegin könnte vielleicht den gleichen Mustern folgen wie unser eigenes („Ach, ich erinnere mich wie niedergeschlagen ich war, als ich bei der Beförderung abgelehnt wurde – ich weiß, wie es dir geht!“). Wenn die Kollegin der ‚Diagnose‘ zustimmt, werden die Dinge noch schwieriger, weil die Zustimmung nicht notwendigerweise bedeutet, dass die Diagnose auch richtig sei. Menschen können überzeugt werden, sich an die Sichtweise anderer, die sie wertschätzen anzupassen. Und auch wenn die Kollegin der ‚Diagnose‘ vehement widerspricht muss das nicht bedeuten, dass sie falsch wäre.

Wenn es um Erziehung geht, kommt man kaum ohne ‚Verstehen‘ aus. Ein etwas ausführlicheres Beispiel mag dazu beitragen, dieses Argument deutlich zu machen. Stellen wir uns eine Dozentin vor, die eine Lehrveranstaltung für Bachelor-Studierenden zum Begriff der Erziehung durchführt. Sie möchte, dass die Studierenden sehen, dass es bei der Erziehung mehr gibt als Examina und Tests zu bestehen und gute Noten auf schriftliche Arbeiten zu bekommen; dass es eine Tradition, besser gesagt: eine Anzahl sich überlappender Traditionen gibt, bei denen es darum geht durch Erziehung Menschen zu ermutigen, ihren eigenen Verstand zu gebrauchen, ihre Horizonte zu erweitern, ihnen dabei zu helfen durchzublicken und sich von den aktuellen Zwängen und Ideologien zu emanzipieren. Zu allen diesen Zielen kommt für eine Dozentin in England gegenwärtig eine weitere Dimension, da Studierende zurzeit erhebliche Kredite aufnehmen müssen, um die Universität besuchen zu können. Die Regierung fordert sie auf, sich als Konsumenten zu definieren, in der Hoffnung, dass sie dann Ansprüche an ihre Dozenten stellen, die wiederum die Standards der universitären Lehre anheben. Die Dozentin möchte, dass die Studierenden einige dieser Absurditäten durchschauen – insbesondere die ganze neoliberale Rhetorik des ‚Marktes‘ als Kontext. Sie möchte durchaus, dass Studierende Ansprüche stellen, aber doch die richtigen. Wenn die Dozentin herausbekommen will ob die Studierenden dies verstanden haben, kann sie sich kaum auf einfache Tests verlassen, teils deswegen, weil die dazugehörigen Begriffe und Gedanken zu komplex sind, aber teils auch wegen der Paradoxie, dass man die Kritik an Tests nicht durch einen Test prüfen kann. Aus diesen und anderen Gründen kann sie deshalb nicht auf gleiche Weise bestimmen, ob die Studierenden das Ziel des Kurses erreicht haben, wie ein Mediziner, der feststellt, dass eine bestimmte Medizin gegen Kopfschmerz geholfen hat. Dies auch deswegen, weil unsere Dozentin auf eigenartige Weise ihren Studierenden Kopfschmerzen zufügen möchte. Wenn ein Studierender nun am Ende der Vorlesung – unter Anführung von z.B. vier guten Gründen – argumentieren würde, dass er vollkommen klar verstanden hätte, dass das Gut der Erziehung nicht wie eine Ware für einen Kunden betrachtet werden könne, würde sich unsere Dozentin die Frage stellen müssen, ob der Studierende nicht schlicht Argumente aus ihrer Vorlesung wiedergibt und sie als dogmatische Sätze behandelt ohne darüber nachzudenken. Vor diesem Hintergrund wäre es auch keine gute Idee, zum Beispiel auf der Grundlage eines kurzen

Fragebogens, zu untersuchen, ob die Studierenden mit ihrer Lehre ‚zufrieden‘ gewesen seien, weil es ja darum ging, sie sowohl zufrieden („Ja, das ist sehr interessant!“) als auch unzufrieden („Ich denke, ihr Argument stimmt nicht, denn in der Vorlesung haben sie gesagt...“) zu machen.

Wie kann diese Dozentin dann überhaupt untersuchen, ob ihre Lehre irgendwie gut gewesen sei? Sie muss die Antworten und Rückmeldungen der Studierenden interpretieren; sie muss verstehen, was die Antworten bedeuten oder worauf sie hinauslaufen. Der Kommentar „Ich denke, ihr Argument stimmt nicht, denn in der Vorlesung haben sie gesagt...“ könnte mit aggressiv-beleidigendem Ton von einer Studentin vorgetragen worden sein, die einen elementaren Widerspruch entdeckt hat und dies als einen Indikator dafür interpretiert, dass die Dozentin die Vorlesung nicht ordentlich vorbereitet habe. Er könnte aber auch mit erregt-begeistertem Ton von einer Studentin formuliert worden sein, die gerade ein Gespür über ihre Fähigkeiten bekommen hat, selbstständig zu denken. Die Studierende, die darüber klagt, dass die empfohlene Lektüre zu schwer sei, mag die Dozentin veranlassen, darüber nachzudenken ob sie nicht einen für Anfängerinnen zu schweren Text gewählt habe. Auf der anderen Seite aber könnte sie freundlich darauf hinweisen, dass der Inhalt dieses Textes – ähnlich wie bei anderen komplexen wissenschaftlichen Texten – sich nicht aus einer einmaligen und vielleicht nur kursorischen Lektüre erschließt. Oder sie wird – erneut im Geist des Verstehens – darauf hinweisen, dass man nicht davon ausgehen kann, dass es nur eine einzige Lesart gebe. Die Dozentin interpretiert die aktuelle Rückfrage einer Studierenden im Lichte der Rückfragen anderer Studierender in Vergangenheit und Gegenwart: Sie ist ihr schon in Veranstaltungen als ernsthafte Person aufgefallen, oder als eine Person, der gesagt werden muss, dass sie nach Hause gehen und den Text ordentlich lesen solle. Möglicherweise wird sie ihr Urteil über die Studierende später revidieren, ähnlich wie sie ihr eigenes Antwortverhalten berücksichtigen und feststellen wird, dass sie zu nachgiebig oder zu ungeduldig oder zu brüsk gewesen sei.

Man kann mit sehr guten Gründen sagen, dass die Dozentin und die Studierende sich gegenseitig lesen, ähnlich wie es im Englischen heißt: „how did you read his behaviour at that meeting?“, wie siehst du das, wie hast du das Verhalten interpretiert? Dies ist kein schnelles und flüchtiges Lesen. Geduld ist ebenso nötig wie eine kontinuierliche und genaue Prüfung der Qualität des ‚Lesens‘. Die Dozentin muss auf der Hut sein vor verallgemeinernden Zuschreibungen, sondern sie eher ‚desinteressiert‘ als ‚bloß‘ Studierende betrachten. Hier ergibt sich erneut eine Analogie zum Lesen: die gute Leserin widmet sich dem besonderen Gedicht, das vor ihr liegt statt Fantasiebilder über Gedichte im Allgemeinen zu imaginieren und zuzuschreiben. Sie stimmt sich eher auf das Gedicht ein als dass sie ein allgemeines und rigides Modell oder eine Methode zur Interpretation aller Gedichte zur Anwendung brächte. Für Text und Studierende gilt, dass es wichtig ist uns reflektierend selbst einzubringen, dass wir bereit sind, uns persönlich irritieren und überraschen zu lassen. Dies kann durch eine fließende Linie in einem Gedicht geschehen oder durch etwas, was im Seminar gesagt wird. Jeder von uns erinnert sich an Diskussionen darüber, warum Studierende zur Universität gehen sollten und was sie meinten, wozu Universität gut sei. Ein Student sagte leise: „das herauszufinden bin ich an die Universität gegangen“. Diesem Moment und Kommentar muss Raum gegeben werden, um mit der gleichen ruhigen Nachdenklichkeit angesprochen werden zu können, mit der es ausgesprochen wurde.

Die Analogie zwischen dem Lesen von Texten und dem Lesen von Menschen führt uns dazu, über Dimensionen von Interpretation und Verstehen im Bereich der Sozialwissenschaften nachzudenken, und über Qualitäten, über die Simone Weil (1951) sagt, dass die Fähigkeit Menschen oder Texten (oder Gott, in ihrer spezifischen Art zu denken) Aufmerksamkeit zu schenken seltener und wertvoller ist als wir üblicherweise meinen. Geduld, Ruhe, Offenheit, Sensibilität sind solche Dimensionen. Sie führen uns auf einen langen Weg und weg von dem überbestimmten ‚wissenschaftlichen‘ und ‚strengen‘, zuweilen auch unerbittlichen Gesicht, das doch so viel erziehungswissenschaftliche Forschung gegenwärtig charakterisiert; und sie führen uns auch weg von der fieberhaften Eile und Geschäftigkeit, die oft als Substitut für kluges und responsives Denken eingesetzt wird. Es geht auch

um die Kunst, warten zu können – wenn wir gefragt werden was wir von diesem Gedicht oder dieser Person halten: „ich muss es noch einmal lesen bevor ich antworten kann“; „ich kenne sie noch nicht gut genug“.

Wir kommen zurück zu einem Beispiel, mit denen wir begonnen hatten: die Studentin, die empirisch darüber forschen wollte, ob Pfadfindersein dazu beiträgt das Selbstwertgefühl zu entwickeln. Hinter dieser Frage scheint die Annahme zu liegen, dass wir wüssten, was Selbstwertgefühl ist. Wir müssen fragen, wie man dies messen kann und wir müssen eine Beziehung zum Pfadfindersein herstellen. Es gibt freilich noch andere knifflige Fragen, die man zum Selbstwertgefühl stellen könnte. Ist es eine gute Sache? Wenn Leute zu viel davon haben, würde es ihnen mit weniger besser gehen? Bedeutet ein geringes Selbstwertgefühl das gleiche wie Bescheidenheit, und wäre Bescheidenheit eine Qualität, von der wir gerne mehr hätten bei unseren Pfadfindern, unseren Kindern, unseren Freunden oder unseren Vizekanzlern? Ist es das gleiche wie Zurückhaltung? Was halten wir von Menschen, von denen wir annehmen, dass ihr Selbstwertgefühl eigentlich tiefe Unsicherheit verdeckt?

Weitere Fragen drängen sich auf. Ist die Steigerung des Selbstwertgefühls das gleiche wie die Entwicklung des Charakters? Die Vorstellung eines ‚Charakters‘ war vor 100 Jahren im anglophonen Westen ausgesprochen populär und markierte ausgesprochen männliche (und militaristische) Eigenschaften wie Loyalität, Mut und Entschlossenheit. Wenn diese Vorstellung einen solchen spezifischen Hintergrund hat, was heißt das dann für das Selbstwertgefühl? Scheinbar ist der Begriff aus der Welt der Beratung und Therapie entstanden und impliziert so etwas wie eine ‚Wendung nach innen‘, eine Selbst-Versenkung, um herauszufinden, ‚wer man eigentlich ist‘. Wir könnten uns fragen ob es das ist, was wir gegenwärtig wollen, insbesondere im Kontext von Pfadfindern, die stets an einer ‚Wendung nach außen‘, am Zelten, am Lernen wie man Feuer macht, an Hilfsbereitschaft gegenüber anderen orientiert waren. Die entscheidende Frage scheint hier die Bedeutung von Selbstwertgefühl zu sein und wie wir sie unterscheiden von anderen Eigenschaften wie Arroganz und Selbst-Versenkung, mit denen dieser Begriff scheinbar leicht zu verwechseln ist.

Solche Fragen nach der Bedeutung führen uns eindeutig in die philosophische Richtung. Sie könnten uns zu den Pfadfinderinnen von Australien führen, die kürzlich ihre Eidesformel geändert haben. Während sie bisher versprochen haben der Queen (d.h. Elisabeth II, der Königin von Großbritannien, dem Vereinigten Königreich und dem Commonwealth) zu dienen und ihre Pflichten gegenüber Gott zu erfüllen, erklären Sie jetzt: „I promise that I will do my best; to be true to myself and develop my beliefs; to serve my community and Australia, and live by the Guide Law“. Der Begriff des Dienens ist erkennbar verbunden mit dem, was Simone Weil mit dem ‚Warten‘ meint. ‚Being true to yourself‘ (was wäre, wenn das ‚Selbst‘, zu dem man ehrlich ist, ziemlich hässlich und böse ist?) und ‚developing your beliefs‘ (was sollte damit gemeint sein?) haben andere Kontexte und Konnotationen. Dies ist eine kleine, aber lehrreiche Illustration, wie uns Fragen über die Entwicklung und die Erziehung junger Menschen eher zu philosophischen Fragen über Bedeutung und Sinn führen als zu ‚wissenschaftlichen‘ Fragen oder empirischer erziehungswissenschaftlicher Forschung.

Fragen über Sinn und Bedeutungen führen uns jedoch auch in die Geschichte von Erziehung und Erziehungswissenschaft, wo ‚Erklären‘ und ‚Verstehen‘ nicht notwendigerweise als Widerspruch interpretiert werden müssen. Beide Begriffe markieren eher unterschiedliche Paradigma, die die Kompetenz des Forschenden herausfordern, die Perspektiven wechseln zu können. Historische Untersuchungen bedienen sich, nutzen und verweisen stets auf beide Paradigmata. Wir wollen sowohl die Strukturen und die Stabilität langdauernder gesellschaftlicher und pädagogischer Prozesse und Mechanismen erklären, wollen aber auch den einzelnen Fall oder die abweichende Gruppe verstehen können – unabhängig davon ob diese mit unseren Erklärungsversuchen übereinstimmen oder nicht. Wie können wir das singuläre ‚Ereignis‘ innerhalb einer konsistenten Theorie der historischen Entwicklung erklären? Wie können wir unsere linearen Erklärungsversuche verbinden mit Erfahrungen einer nicht-linearen Komplexität von Formen, Aspekten und Ebenen des Verstehens?

Wie können wir das Vergleichbare vom Nicht-vergleichbaren oder vom Unvergleichlichen unterscheiden (Depaepe 2002)?

Erziehungswissenschaft ist in dieser Hinsicht ein interessanter und herausfordernder Gegenstand, der sich besonders gut zur Forschung über Forschung eignet – in empirischer, analytischer, historischer und vergleichender Perspektive. Erziehungswissenschaft ist in außerordentlich hohem Maße auf Paradoxien, Widersprüche und Antinomien gegründet und in diese eingebettet, die wir gegen alle Standardisierungsversuche bewusst halten sollten. Wir sollten diese in das Fach ‚eingebauten‘ Paradoxien als systematische Spannungen und Beziehungen nutzen, insbesondere um schwierige und aufregende Fragen gegen vorschnelle und wohlfeile Antworten zu verteidigen.

Literatur

- Ambrose, D. (2006). Large-scale contextual influences on creativity: Evolving academic disciplines and global value systems. *Creativity Research Journal* 18, 75-85.
- Bourdieu, P. (1998). *Vom Gebrauch der Wissenschaft. Für eine klinische Soziologie des wissenschaftlichen Feldes*. Universitätsverlag Konstanz: Konstanz.
- Burbules, N. (2013). Learning is not education. In P. Smeyers & M. Depaepe (eds), *Educational Research: The Attraction of Psychology* (S. 159-166). Dordrecht: Springer.
- Cartwright, N. & Hardie, J. (2012). *Evidence-based Policy*. Oxford: Oxford University Press.
- Depaepe, M. (2002). A Comparative History of Educational Sciences: the comparability of the incomparable? *European Educational Research Journal* 1(1), 118-122.
- Depaepe, M. (2010). Dazzling statistics? On the university expansion in Flanders and the need for research into the history of education that transcends quantifying sociology. In P. Smeyers & M. Depaepe (eds), *Educational Research: The Ethics and Aesthetics of Statistics* (S. 27-42). Dordrecht: Springer.
- Fatke, R. & Oelkers, J. (Hrsg.) (2014). *Das Selbstverständnis der Erziehungswissenschaft: Geschichte und Gegenwart* (60. Beiheft der Zeitschrift für Pädagogik). Weinheim, Basel: Beltz.
- Hargreaves, D. (1996). *Teaching as a Research-based Profession: possibilities and prospects. The Teacher Training Agency Annual Lecture*. London: The Teacher Training Agency
- Hodgson, N. & Standish, P. (2007). Network, critique, conversation: towards a rethinking of educational research methods training. In P. Smeyers & M. Depaepe (eds). *Educational Research: Networks and Technologies* (S. 105-120). Dordrecht: Springer.
- Hoffmann, D. & Neumann, K. (Hrsg.) (1998). *Die gegenwärtige Struktur der Erziehungswissenschaft. Zum Selbstverständnis einer undisziplinierten Disziplin*. Weinheim: Deutscher Studienverlag.
- Keiner, E. & Schaufler, S. (2014). Disziplinäre und organisatorische Grenzen, Überschneidungen und Neuformatierungen – Das Beispiel Pädagogische Psychologie und Erziehungswissenschaft. In N. Ricken, H.-C. Koller & E. Keiner (Hrsg.), *Die Idee der Universität – Revisited* (S. 269-301). Wiesbaden: Springer VS.
- Keiner, E. & Schriewer, J. (2000). Erneuerung aus dem Geist der eigenen Tradition? Über Kontinuität und Wandel nationaler Denkstile in der Erziehungswissenschaft. *Schweizerische Zeitschrift für Bildungswissenschaften/Revue suisse des sciences de l'éducation/Rivista svizzera di scienze dell'educazione* 22(1), 27-50.
- Keiner, E. (2006). The science of education – disciplinary knowledge or non-knowledge/ignorance? In P. Smeyers and M. Depaepe (eds), *Educational Research: why 'what works' doesn't work* (S. 171-186). Dordrecht: Springer.
- Keiner, E. (2013). The Short Timelines of History of Education at Present. *International Journal of the History of Education*, 2, 234-237.
- Keiner, E. (2015). Pädagogik, Erziehungswissenschaft, Bildungswissenschaft, Empirische Bildungsforschung – Begriffe und funktionale Kontexte. In E. Glaser & E. Keiner (Hrsg.), *Unscharfe Grenzen – eine Disziplin im Dialog: Pädagogik, Erziehungswissenschaft, Bildungswissenschaft, Empirische Bildungsforschung* (S. 13-34). Bad Heilbrunn: Klinkhardt.
- Knaupp, M., Schaufler, S., Hofbauer, S. & Keiner, E. (2014). Education Research and Educational Psychology in Germany, Italy and the United Kingdom – an Analysis of Scholarly Journals. *Schweizerische Zeitschrift für Bildungswissenschaften/Revue suisse des sciences de l'éducation/Rivista svizzera di scienze dell'educazione*, 36, 83-106.
- Meusburger, P. (2009). Räumliche Disparitäten des Wissens. Zu einigen Kommunikationsdefiziten zwischen wissenschaftlichen Disziplinen. In M. Hey & K. Engert (Hrsg.). *Komplexe Regionen –*

- Regionenkomplexe. Multiperspektivische Ansätze zur Beschreibung regionaler und urbaner Dynamiken* (S. 209-229). Wiesbaden: VS-Verlag.
- Midgley, M. (1996). Philosophical plumbing. In *Utopias, Dolphins and Computers* (S. 1-14). London: Routledge.
- Nóvoa, A. & Yariv-Marshall, T. (2003). Comparative research in education: a mode of governance or a historical journey? *Comparative Education*, 39(4), 423-438.
- Oancea A. & Pring, R. (2008). The importance of being thorough: on systematic accumulations of 'what works' in education research. *Journal of Philosophy of Education*, 42(1), 4-15.
- Smeyers P. & Depaepe M. (eds) (2003). *Beyond Empiricism. On Criteria for Educational Research*. Leuven: Leuven University Press.
- Smeyers P. & Depaepe M. (eds) (2006). *Educational Research: Why 'What Works' Doesn't Work*. Dordrecht: Springer.
- Smeyers, P. (2006). The relevance of irrelevant research; the irrelevance of relevant research. In P. Smeyers & M. Depaepe (eds), *Educational Research: why 'what works' doesn't work* (S. 95-108). Dordrecht: Springer.
- Smeyers, P. (2007). The hidden homogenisation of educational research: on opening up the sphere of responsiveness. In P. Smeyers & M. Depaepe (eds), *Educational Research: Networks and Technologies* (S. 85-104). Dordrecht: Springer.
- Smeyers, P. & Depaepe, M. (eds) (2010). *Educational research: The ethics and aesthetics of statistics*. Dordrecht: Springer.
- Smeyers, P. & Depaepe, M. (eds.) (2008). *Educational research: The Educationalization of social problems*. Dordrecht: Springer.
- Smeyers, P. & Depaepe, M. (eds.) (2013). *Educational research: The attraction of psychology*. Dordrecht: Springer.
- Smith, R. (2010). A bubble for the spirit level: metricophilia, rhetoric and philosophy. In P. Smeyers & M. Depaepe (eds), *Educational Research: the ethics and aesthetics of statistics* (S. 189-204). Dordrecht: Springer.
- Stadler-Altman, U. & Keiner, E. (2010). The Persuasive Power of Figures and the Aesthetics of the Dirty Backyards of Statistics in Educational Research. In P. Smeyers & M. Depaepe (eds), *Educational Research - the Ethics and Aesthetics of Statistics* (S. 129-144). Dordrecht: Springer.
- Tenorth, H.-E. (1983). Die Krisen der Theoretiker sind nicht die Krisen der Theorie. *Zeitschrift für Pädagogik*, 29, 347-358.
- Tooley, J. (1998). *Educational Research: a critique*. London: Office for Standards in Education.
- Weil, S. (1951). Reflections on the right use of school studies with a view to the love of God. In S. Weil, *Waiting on God* (trans. E. Crauford) (S. 105-116). New York: G.P. Putnam's Sons.
- Wright, G. H. von (1971). *Explanation and Understanding*. London Routledge & Kegan Paul.